

# Effects of County-Level Diversity on Educational, Economic, and Health-Related Outcomes

*Kenny Groszman, Jack Petersen, Bing Xu*

## Introduction:

It is often touted by the government, businesses, and members of our community that diversity brings with it a variety of benefits [1]. Indeed, it has been shown that at the organization-level, diversity does provide micro-scale benefits such as new perspectives, increased productivity, expanded worldviews for the members of the organization, and more [2]. This effect leads us to ask the following question: *does diversity bring with it macro-scale benefits? More specifically, is being in a more diverse county correlated with statistical indicators of success in a variety of different categories?*

We will consider those three categories of possible downstream effects of diversity: educational, economic, and health-related outcomes. It is possible that increased diversity will actually have negative effects on a broader scale, as certain minority communities may experience alienation. However, it is also possible that different racial groups will live harmoniously, learn from each other, and contribute to an overall more successful community.

It should also be noted that right off the bat, these categories of success that we are looking at are affected by a plethora of factors that do not include racial diversity (geographic location, history, politics, major industries and employers, etc.). Thus, we expect that correlations will be weak. However, we also expect to be able to glean some insights about the effects of broad diversity in America.

## Data

The approach to data collection for this project was to take several different county-level data sets and to join them using the unique Federal Information Processing Standard (FIPS) code for each county. Education, Health, and Economics data were taken from county-level databases published by several different government agencies. However, county diversity and race data were constructed from a household-level database called the Integrated Public Use Microdata Series (IPUMS), which was then grouped by county for further analysis.

The variables we were able to pull from these databases and use for analysis, as well as their sources, are below:

Variable	Description	Source
Geographic Location	County location of record, given by state and county FIPS (Federal Information Processing Standard) code	[3]
Race	Numerical code corresponding to race of individual: white, black, American Indian, Chinese, Japanese, Pacific Islander, other, 2+ races	[3]
Educational Attainment	Code corresponding to highest level of education achieved: no high school, high school diploma, Associate's degree, Bachelor's degree	[4]
Poverty Status	Code indicating whether individual falls below the poverty line	[5]
Income	Integer indicating the household's yearly income	[5]
Life Expectancy	Database of life expectancy for different counties in the US	[6]
Obesity	Rates of obesity in different US counties	[6]

## Creating a Diversity Index

Upon summarizing race data by county, race information is given as the proportion of the county population that falls into 8 different groups:

- White
- Black
- American Indian/Alaskan Native
- Chinese
- Japanese
- Other Asian / Pacific Islander
- Other Race
- 2+ races

We needed a measure of Diversity that follows the following conditions:

1. The county with the highest possible Diversity Index Score is one with the population spread evenly across the 8 groups (i.e. 12.5% in each group)
2. A county with the lowest possible Diversity Index Score is one with the population entirely in one of the groups (i.e. 100% White, 0% everything else)

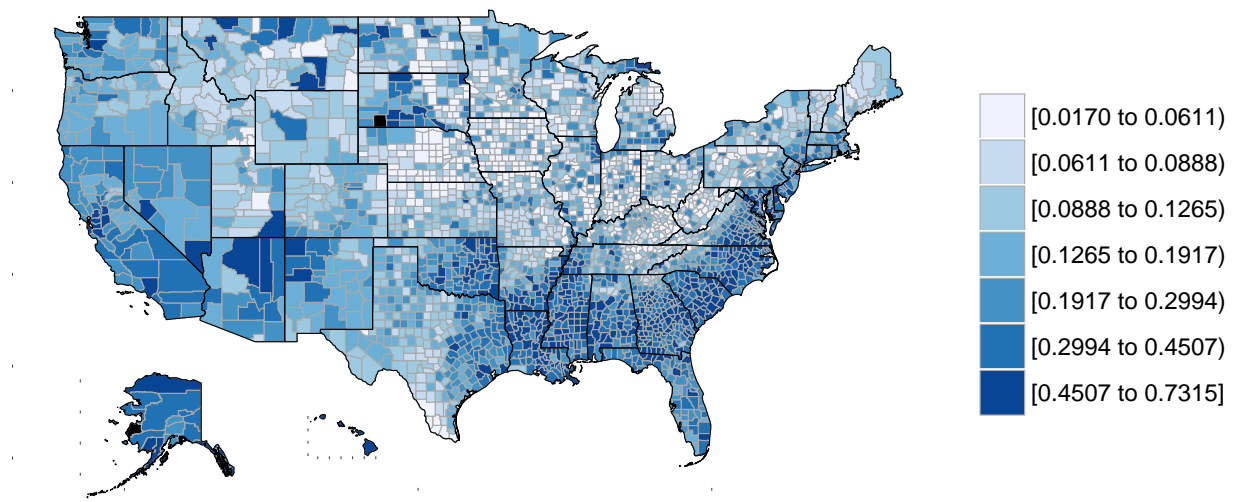
A formula that meets this criteria for county  $i$  is as follows, for races  $R \in \{White, Black, \dots, 2 + Races\}$  :

$$DiversityIndex_i = 1 - \sum_{j \in R} x_{ij}^2$$

where  $x_{ij}$  is the proportion of race  $j$  in county  $i$  [7].

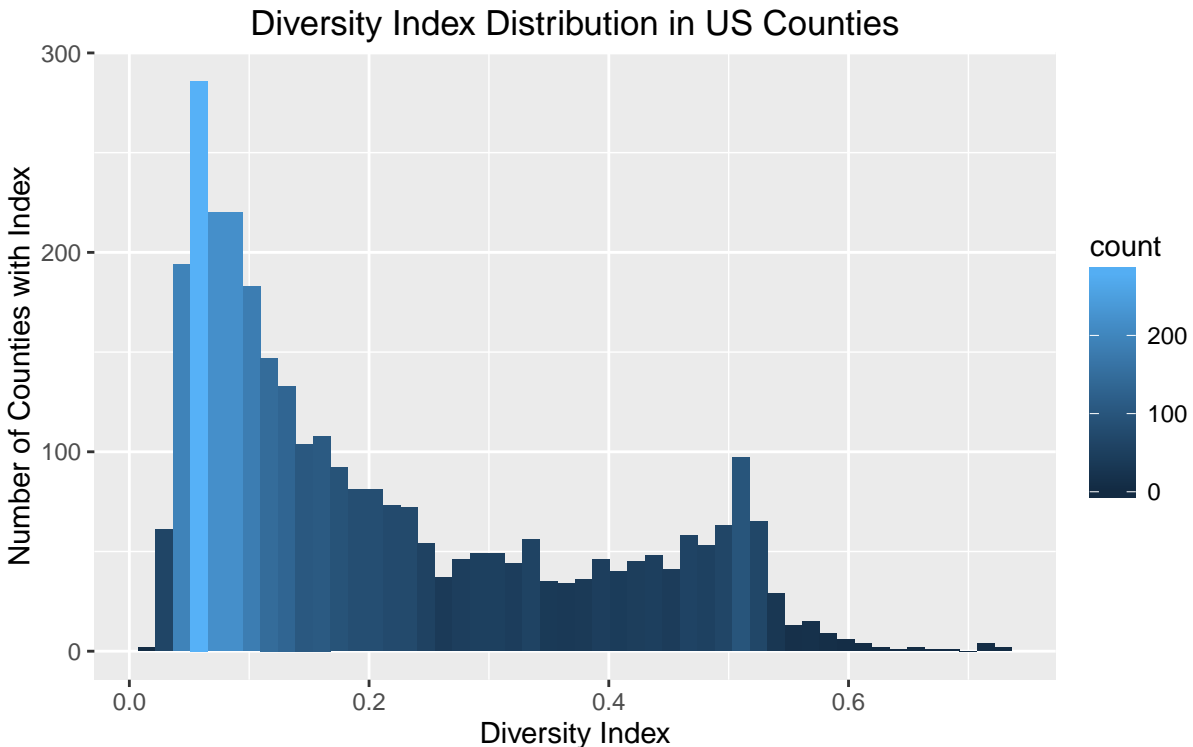
## Results and Interpretation

### Diversity Index in Counties Across the US



We see from this map that diverse counties in the US are concentrated near the country's southern, western, and eastern borders. Additionally, Alaska is quite diverse. To the contrary, the midwest and central USA are among the lowest diversity counties. These results are important to keep in mind as we analyze downstream effects, as particulars about these regions may confound the analysis. For example, if we were to look at hospital quality based on diversity, our results might be skewed since the best hospitals in the country are in California and New York, for reasons that are probably independent of racial diversity.

Another interesting note is that there are far more counties with quite low levels of diversity (0 - 0.2) than there are counties with higher levels (0.2 - 0.6). There does seem to be a spike at around 0.5, and this might be due to large coastal counties. Very few counties have a diversity index greater than 0.6.

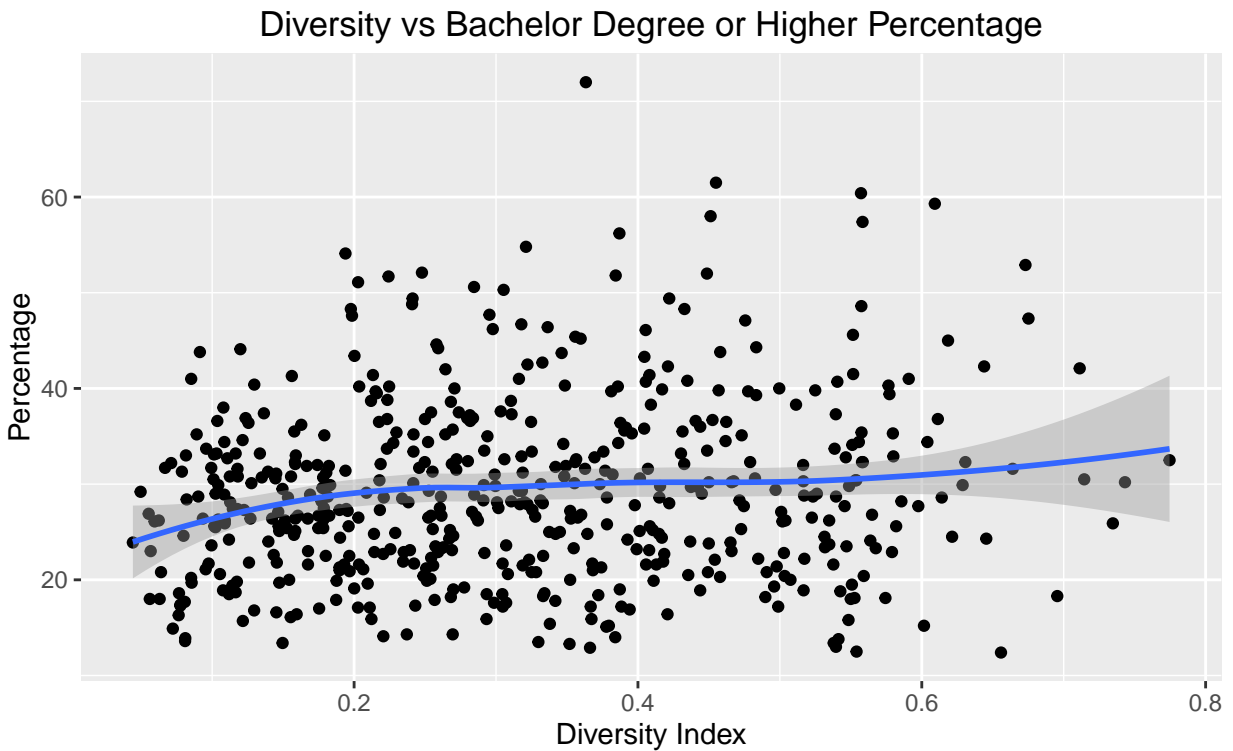
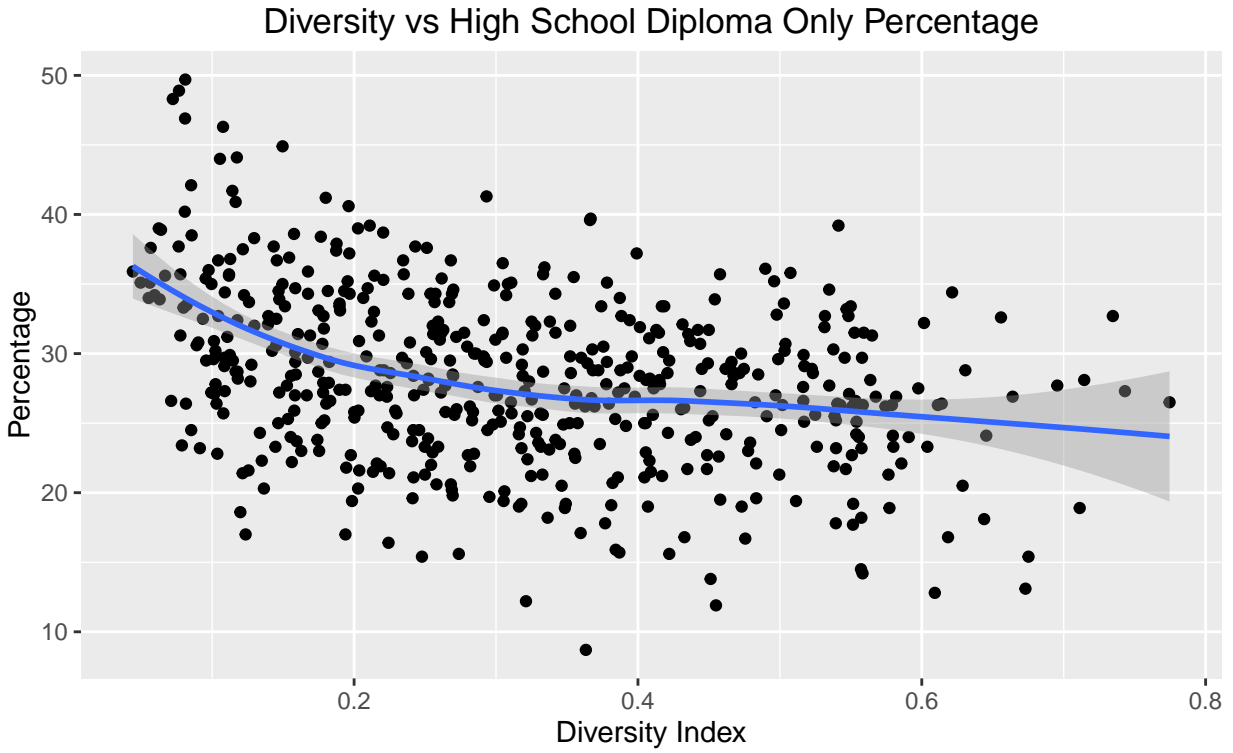


## Effects of Diversity on Educational Outcomes

Here, we pull educational data from various sources and compare it to county-level diversity data. The hope is that we would see a relationship between diversity levels and the educational trends in a county. Here we define educational success as the rates of completion for several different levels of education.

### *Relationship between Diversity and Percentage of Adults with Particular Levels of Educational Attainment*

From the US Department of Agriculture Economic Research Service, we were able to find a data set that includes the percentage of adults attaining certain education levels for each county. So we are interested in finding whether there are relationships between the diversity we devised and percentages of completing degrees. There are four percentages, which are percent of adults with less than a high school diploma, percent of adults with a high school diploma only, percent of adults completing some college or associate's degree, percent of adults with a bachelor's degree or higher.



***Commentary and Interpretation:***

From the above graphs, we see that more diverse counties have a higher percentage of students reaching the highest levels of education. On the other side, the least diverse counties have most number of people with low educational attainment (i.e. only finishing high school).

It is possible that these educational trends are correlated with economic trends or demographic trends. For example, since most counties with a low diversity index are in the middle of the USA, it is possible that students are dropping out of school after high school to do manual labor (i.e. farming, manufacturing, etc.) whereas the more diverse counties are in areas where higher education levels are required.

However, it is also distinctly possible that in more diverse areas, minority groups receive more educational opportunities and thus end up receiving higher education, whereas those same groups in less diverse areas wouldn't (due to marginalization).

## Effects of Diversity on Economic Outcomes

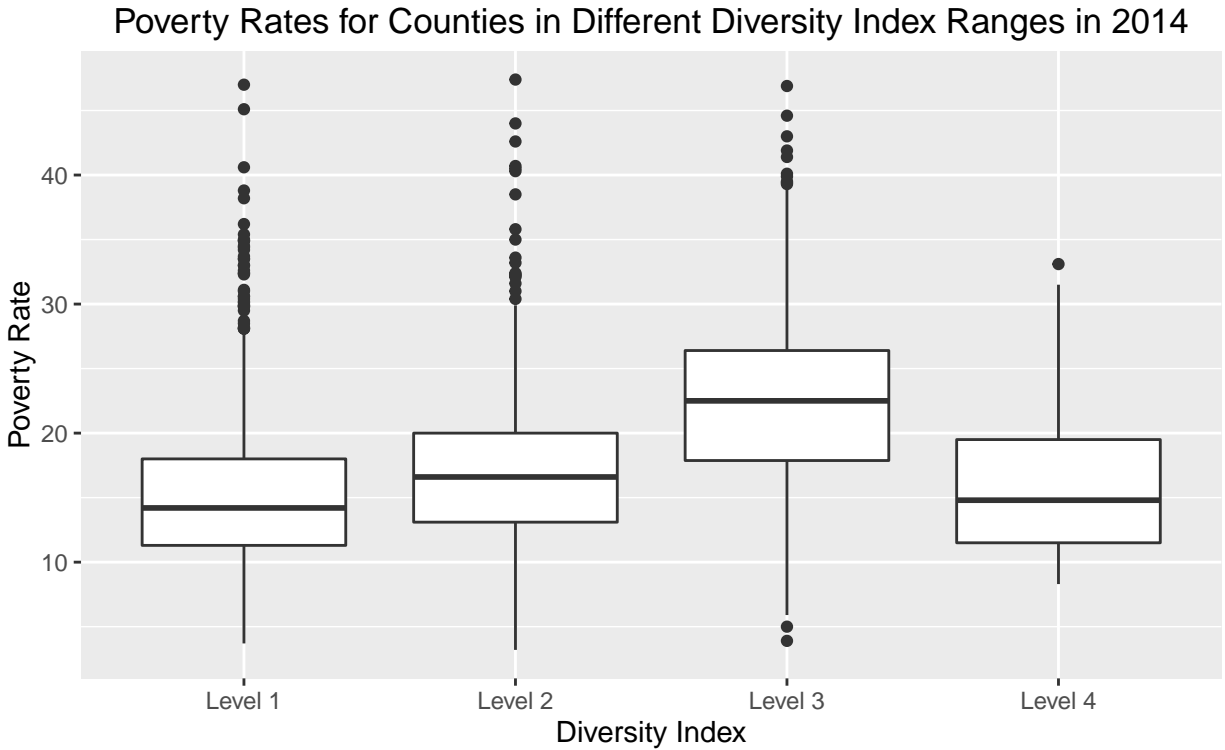
In this section, we are going to look at whether there is a relationship between the diversity and economic outcomes within counties. We are using variables like poverty rate, unemployment rate, median income and the income gap as indicators of counties' economic outcomes.

### *1. US Census Bureau Poverty Estimates*

US Census Bureau Model-based Small Area Income & Poverty Estimates provides the estimates of the total population and the population of falling below the poverty line in each county in 2014 and we get the poverty rate estimates from this dataset. And we categorize counties into four types by the diversity index. The four types are:

- Level 1: The diversity index is from 0 to 0.2.
- Level 2: The diversity index is from 0.2 to 0.4.
- Level 3: The diversity index is from 0.4 to 0.6.
- Level 4: The diversity index is from 0.6 to 0.8.

Then we draw the following graph and try to find whether there are differences of poverty rates in different diversity index ranges.



***Commentary and Interpretation:***

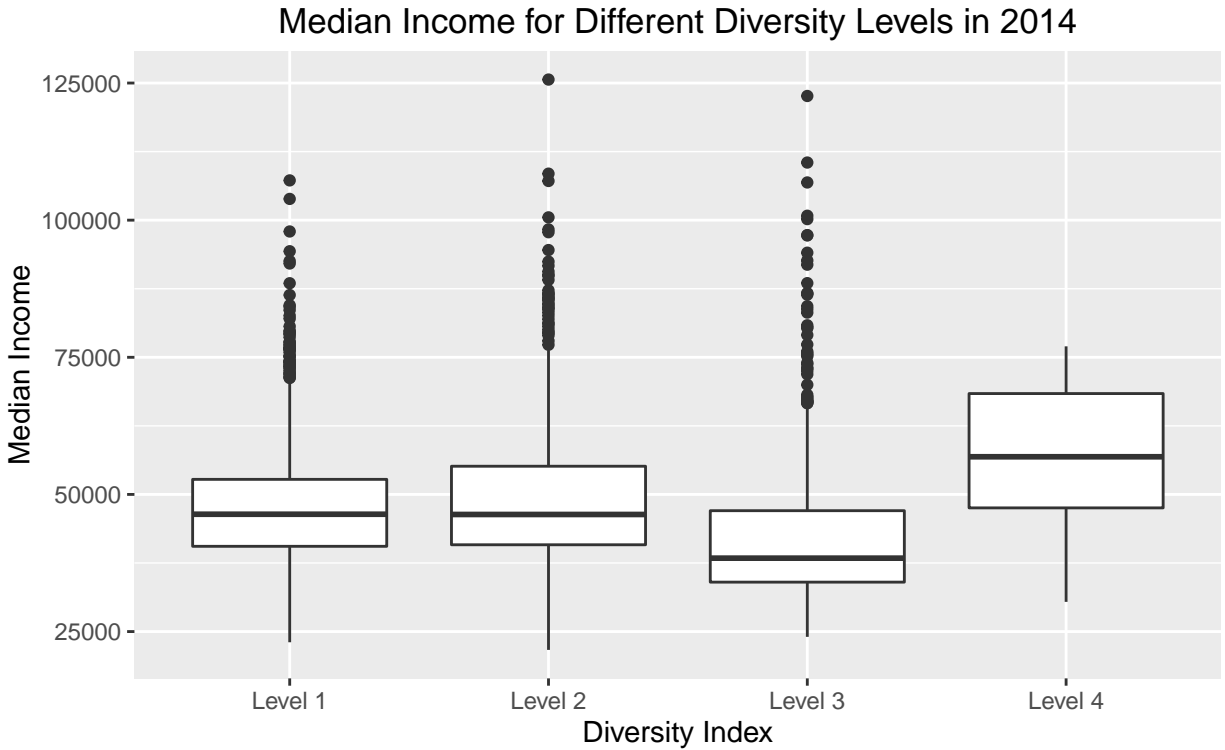
We can observe that when a county's diversity index is extremely low or high, its poverty rate tends to be relatively low. For counties that have an intermediate diversity index, the poverty rates are increasing. The poverty rates of Level 3 are significantly higher than other groups'.

One possible explanation is that in counties with intermediate diversity levels, some minority groups are possibly marginalized and these groups are more easily to be in poverty status which raise the poverty rate. While in slightly or highly diversified counties, this situation will be unlikely to occur.

***2. US Census Bureau Median Household Income***

From US Census Bureau Model-based Small Area Income & Poverty Estimates, we can also get the annual median household income data for each county in 2014.

We create the similar boxplot using the median household income and the diversity index level.



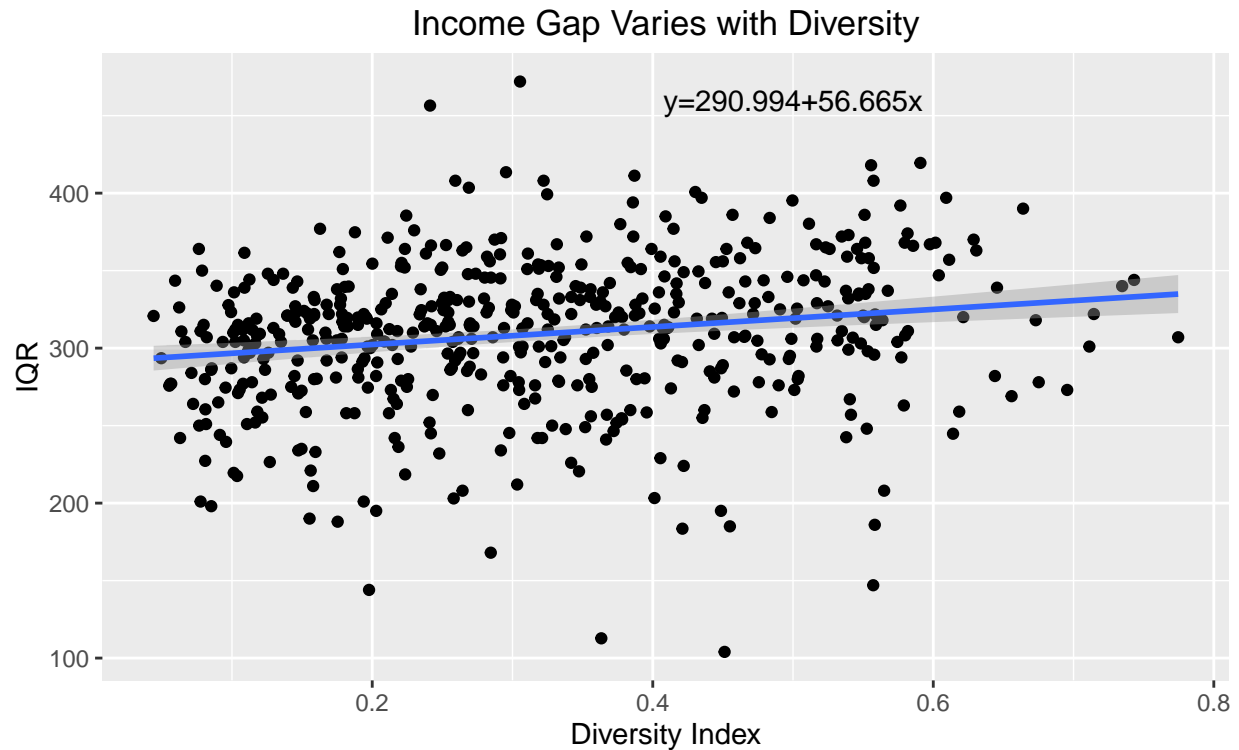
#### *Commentary and Interpretation:*

We can see that the median household income for Level 3 is lower than other groups', which is consistent with results in the poverty rate part. So the low median household income for Level 3 may be due to the marginalization of some minority groups.

We can also notice that Level 4's median household income is the higher than other three groups'. One possible reason is that highly diversified counties are more international and more international regions are more likely to have higher income levels.

### **3. IPUMS Income Gap**

We also use the IPUMS dataset to observe the relationship between the income gap and the diversity index. We use the interquartile range of household incomes for a given county as the income gap.

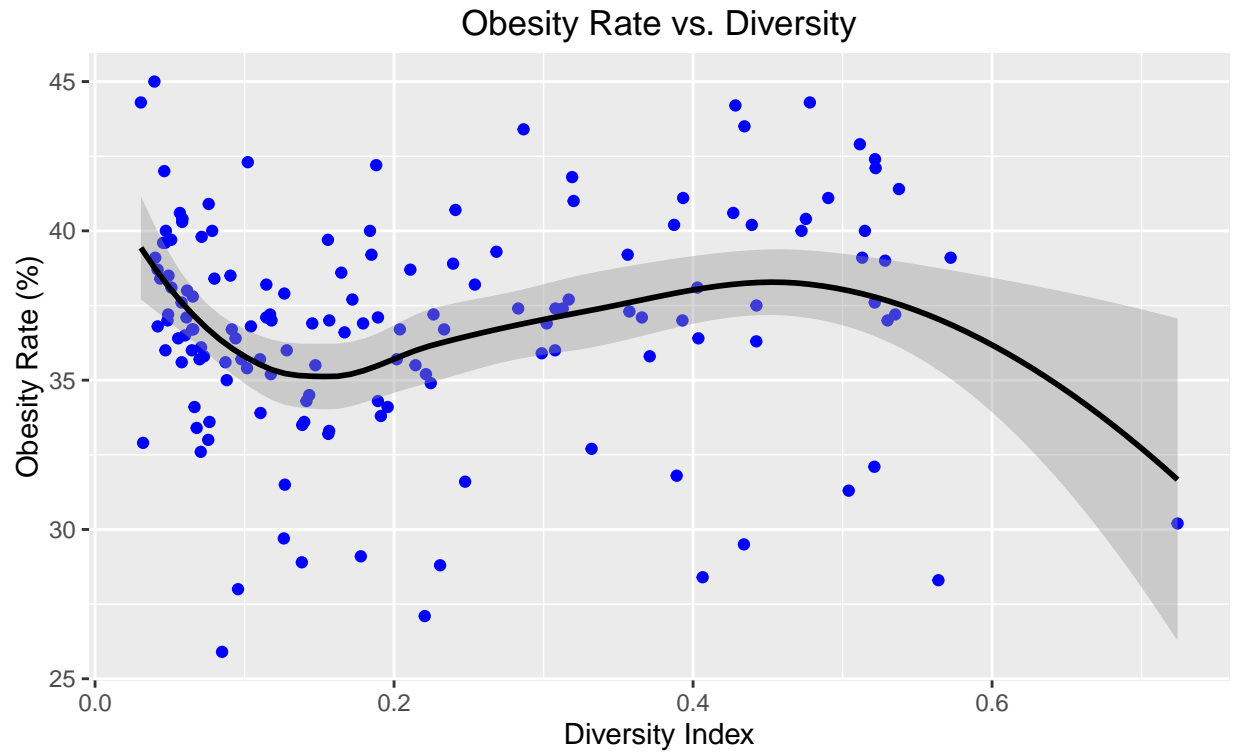


We observe a positive relationship between the income gap and the diversity index from the scatterplot. The correlation is not very strong as the income gap could be affected by many factors. A county with a higher diversity index tends to have a more obvious income gap. This trend may be explained by the racial divide between highest and lowest-paying jobs, which increase the income gap. While for less diverse counties, this phenomenon is not that severe.

## Analyzing Diversity Effects on Health Outcomes

Here, we examine how the diversity of a county plays into the relative health of those living in the county. We used diversity data from an index detailed above, and census data to examine health effects.

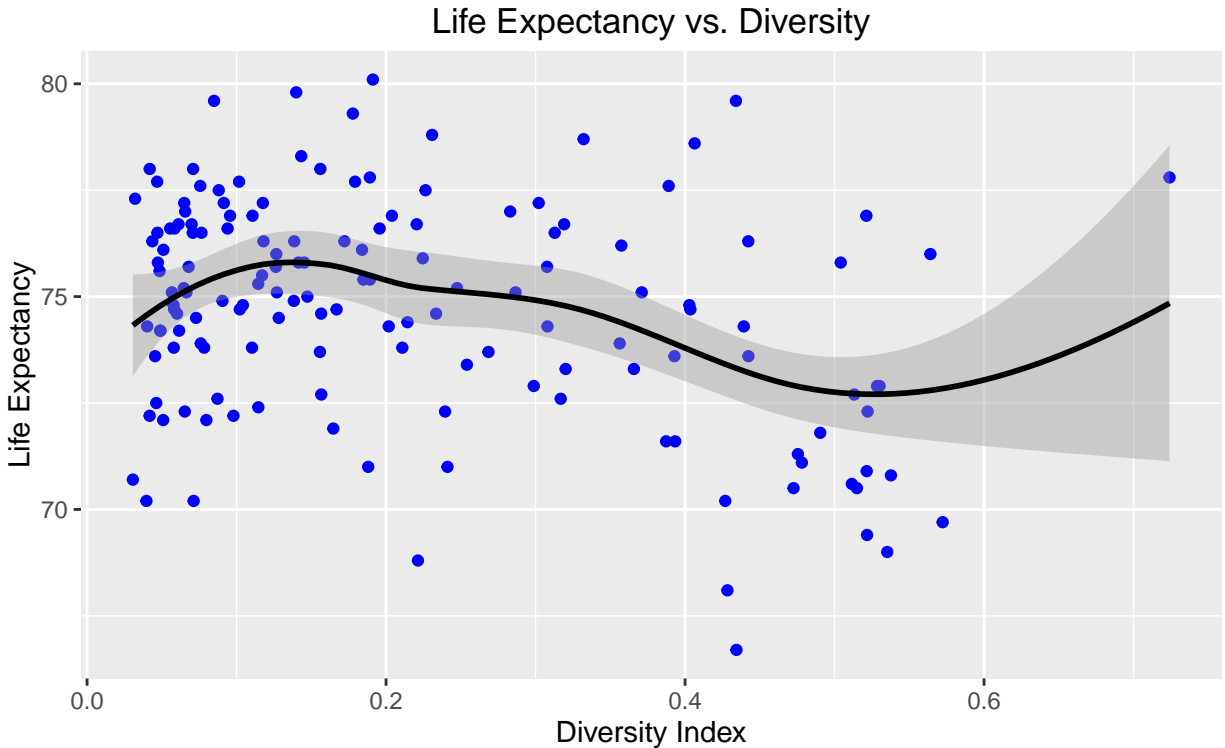




***Commentary and Interpretation:***

From the graph, we see that obesity rates tend to be the highest in counties with medium high diversity. This could be correlated with higher poverty levels and that people in these counties eat more junk food. However, more data is necessary to make a confident conclusion on the matter.

Using the diversity index calculated before, we now look at health outcomes using a set of U.S. Census Data that measures obesity and life expectancy. This was compared against county diversity data.



#### ***Commentary and Interpretation:***

From the graph, we can see that life expectancy decreases with increased diversity. We came up with a few possible causes for this. First, a more diverse county is likely to be urban and have a higher risk of communicable disease. Additionally, certain racial groups have lower life expectancies which would drive down the life expectancy of more urban counties

### **Future Plans and Conclusions**

#### **Future Plans**

As always, there is more analysis that could be done that was outside the scope of this project. For example, it would be useful to control for racial differences in outcomes (for example, differences in obesity rates in white vs. black Americans). Thus, it would be helpful to look at the different performance of the different races in counties with different diversity levels. It might also be useful to look at this data in the reverse order. For example, we could ask the question “do people with higher incomes tend to be in more or less diverse counties?”. This might give us some additional insights about causation that our current analysis is lacking. With more time and more data, both of these research directions would provide extremely interesting results.

#### **Conclusion**

In all, county-level diversity is indeed correlated with certain measures of educational, economic, and health-related outcomes. However, the analysis is quite limited since many other external factors also play into these outcomes (size of county, location, major industries, urbanization status, etc.). In general, we see a certain tradeoff between diversity and beneficial outcomes. For example, in the economic section, we saw that intermediate levels of diversity were associated with higher poverty rates and lower income, and we posit that this might be caused by certain racial groups being marginalized in these counties. This is bolstered by the

observation that at the highest diversity levels, the poverty rates go back down. This might be because if a certain minority group is present in high enough proportions, this marginalization does not occur. Similarly, obesity is also highest in intermediate-diversity counties.

This leads us to hypothesize that living in an extremely undiverse or an extremely diverse county results in better outcomes, and this might just be because in these counties, one has a community of people very similar to themselves on which to rely on and feel comfortable with. To the contrary, in counties with many different groups present in small numbers, outcomes are generally worse.

With that, it looks like the groups who tout the categorical benefits of diversity are only partially correct.

## References

- [1] Are there really any tangible benefits to diversity and inclusion? Isn't this just corporate jargon? US office of personnel management, (n.d.). <https://www.opm.gov/faqs/QA.aspx?fid=72bcd219-0b9f-4de8-b366-4817028fbc6e&pid=f2ef3151-b4f2-4f47-a319-acad8175b0b7> (accessed November 28, 2016).
- [2] K. Abreu, The myriad benefits of diversity in the workplace. Entrepreneur, (2014). <https://www.entrepreneur.com/article/240550> (accessed November 28, 2016).
- [3] S. Ruggles, K. Genadek, R. Goeken, M. Sobek, Integrated public use microdata series: Version 6.0 [machine-readable database], University of Minnesota, Minneapolis, 2015. <https://usa.ipums.org/usa/> (accessed November 30, 2016).
- [4] 2015 US county race data. US census bureau, (n.d.). <https://www.census.gov/popest/data/counties/asrh/2015/files/CC-EST2015-ALLDATA.csv> (accessed November 30, 2016).
- [5] USDA-ERS county-level datasets. United states department of agriculture economic research service, (n.d.). <https://www.ers.usda.gov/data-products/county-level-data-sets/download-data.aspx> (accessed November 30, 2016).
- [6] IHME US data for download. Institute for health metrics and evaluation, (n.d.). <http://www.healthdata.org/us-health/data-download> (accessed November 30, 2016).
- [7] Diversity index of US counties kaggle, (n.d.). <https://www.kaggle.com/mikejohnsonjr/us-counties-diversity-index> (accessed November 30, 2016).